

Cardiac Stroke Prediction Framework using Hybrid Optimization Algorithm under DNN

Naif K. Al-Shammari

Mechanical Engineering Department
College of Engineering
University of Ha'il
Ha'il, Saudi Arabia
naif.alshammari@uoh.edu.sa

Ahmed A. Alzamil

Electrical Engineering Department
College of Engineering
University of Ha'il
Ha'il, Saudi Arabia
aa.alzamil@uoh.edu.sa

Mohammed Albadarn

Electrical Engineering Department
College of Engineering
University of Ha'il
H'ail, Saudi Arabia
s.abadran@uoh.edu.sa

Syed Thouheed Ahmed

School of Computing and Information
Technology
REVA University
Bangaluru, Karnataka, India
syedthouheed.ahmed@reva.edu.in

Syed Muzamil Basha

School of Computer Science and
Engineering
REVA University
Bangaluru, Karnataka, India
muzamilbasha.s@reva.edu.in

Ahmed Saud Alshammari

Electrical Engineering Department
College of Engineering
University of Ha'il
Ha'il, Saudi Arabia
ahm.alshammari@uoh.edu.sa

Ahmed Maher Gabr

Physical Therapy Department
Faculty of Applied Medical Sciences
University of Ha'il
Ha'il, Saudi Arabia
am.mohammed@uoh.edu.sa

Abstract-Heart weakness and restricted blood flow into the cavities can cause a range of strokes from mild to severe Heart strokes are primary caused due to the fat deposited on artery walls. The process reduces the intake of blood and internally causes a pseudo vacuum of air bubbles leading to a stroke which can be identified with high-end instrumentations. In this article, a detailed evaluation is processed with a Hybrid Optimization Algorithm (HOA). In the proposed technique, data are preprocessed using a label encoder and the missing values of the dataset are filled. Whale Optimization Algorithm (WOA) and Crow Search Algorithm(CSA) extract inter-connected patterns and learning features using a dedicated Deep Neural Networking (DNN) support. The proposed Hybrid Optimization Algorithm extracts features and the resultant values demonstrate a high accuracy range of 97.34%.

Keywords-cardiac stroke prediction; cardiovascular disease; whale optimization algorithm; crow search algorithm; hybrid optimization algorithm

I. INTRODUCTION

A cardiac stroke occurs when the blood flow to the heart is blocked. A buildup of fat, saturated fat, and other substances in the blood vessel that stream the heart, known as plaque, is the most prevalent cause of obstruction. Plaque can break and generate clot, blocking blood flow. Some portions of the heart muscle can get injured or damaged if the blood flow is

disrupted. Some of the common symptoms of a cardiac stroke are pressure, nausea, heartburn, fatigue, and shortness of breath. Not all patients have the same symptoms or symptom intensity. Some people experience modest pain, while others severe agony [1]. Some persons have no signs or symptoms. For others, abrupt cardiac arrest may be the first symptom. The larger the number of symptoms one experiences, the more likely the patient suffers a heart attack. Although some heart attacks occur unexpectedly, many patients have warning signs and symptoms such as recurrent chest pain that is provoked by exercise and eased by rest.

According to The World Heart Federation, by 2030 there would be more than 23 million heart disease-related deaths annually [2]. Heart related diseases are the most common cause of death in the US according to the American Heart Association. In the US, over 605,000 new congestive heart failures and 200,000 frequent congestive heart failures take place each year [3]. The process of predicting a stroke is commonly a cardiologist's task. These experts rely on medical information and practical demographic images such as Computerized Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET), invalidating the blockages. Still, due to the more significant extent of the population and the lack of sufficient number of experts, the validation process is challenging and with a high

Corresponding author: Naif Khalaf Al-Shammari

www.etasr.com

Al-Shammari et al.: Cardiac Stroke Prediction Framework using Hybrid Optimization Algorithm ...

order of false-positive predictions. The cardiac stroke is validated under higher-computational techniques to resolve an exact process of prediction by experts. The computational approach improves a prediction to validation ratio as it deals with sensitive medical information processing [4, 25]. The association of Deep Neural Networks (DNNs) provides a reliable process for data mapping and digitalization-based learning. The learned datasets are pre-processed and synchronized timely to form reliable training datasets. These datasets can be used for primary filtering and for providing accurate decision support. The inclusion of deep learning datasets improves the prediction rate and the dependability ratio.

II. LITERATURE REVIEW

Cardiac strokes are also known as heart attacks, heart jamming, vascular blockages, etc. Predicting a heart stroke is a challenging endeavor. Fuzzy Clustering Means (FCM), an unsupervised classification technique, was used to envisage early heart attacks on utilizing the patient medical information. Data mining techniques were used to preprocess the information in the patient record, and a Fuzzy C means classifier was used to classify the attributes in [5]. The effectiveness of the classifier was evaluated using data from 270 patients and had an accuracy of 92%. Medical professionals can use a cardiac stroke prediction system to forecast patients' heart disease status based on their clinical data [6]. Heart disease was classified using an artificial neural network algorithm, with 80% accuracy [6]. To assess the chance of increase in heart disease, the algorithm considers 13 factors. Heart disease database tests classification algorithms such as Decision Trees, Naive Bayes, and Neural Networks were utilized in [7] with 100%, 99.62%, and 90.74% accuracy for Neural Networks, Decision Trees, and Naive Bayes respectively. A Neural Network was merged in a coactive neuro-fuzzy inference system in [8]. The presence of the disease was diagnosed with a fuzzy logic qualitative technique combined with a genetic algorithm. The Coactive Neuro-Fuzzy Inference System (CANFIS) performance was assessed in terms of training results with 0.000842 mean square error. The system for predicting heart disease in [9] used ensemble deep learning and feature fusion methods. The information gain methodology reduced the computational burden and improved system performance by removing inappropriate and unnecessary components with 98.5% accuracy.

The enhanced deep learning aided convolutional neural network in [10] increased cardiac stroke prediction. It combines multi-layer perceptron's model with learning techniques and was executed on an Internet of Medical Things platform. It achieved a precision as high as 99.1%. The Hybrid Random Forest with a Linear Model (HRFLM) tried to uncover relevant features using machine learning approaches, enhancing the accuracy of cardiac stroke estimation with various combinations of data [11] with an accuracy of 88.7%. To distinguish between heart disease patients and healthy people, researchers in [12] employed a deep learning system based on Multiple Kernel Learning (MLK) with Adaptive Neuro-Fuzzy Inference System (ANFIS). The ANFIS classifier uses the MLK approach to categorize heart disease and healthy

subjects. The system offers 98% sensitivity, 99% specificity, and 0.01% mean square error. For the initialization of neural network weights, the hybrid system in [13] employs a global optimization genetic method. The learning is faster, more consistent, and accurate, with an accuracy of 89% in predicting the risk of heart disease.

III. METHODOLOGY

The dataset for the proposed system is taken from the open Cleveland data base repository of 270 patients [14]. This multi-modal dataset needed to be pre-processed before it could be used in the deep learning model. The misplaced values in the dataset were filled in the pre-processing step. Using the Label Encoder approach, the data were converted to a numerical representation. Resampling technique was used to reduce data imbalances, and the standard scalar methodology was used to normalize the data to a value ranging from 0 to 1. In the proposed system, we used 13 attributes for feature extraction: age, gender, chest pain type, blood pressure, cholesterol, blood sugar during fasting, rest ECG, thal (3 = normal; 6 = fixed defect; 7 = reversible defect), exercise-induced angina, old peak (ST depression induced by exercise relative to rest), slope of the peak exercise ST segment, maximum heart rate, and number of major vessels colored by fluoroscopy. There are significant and insignificant features in this normalized dataset. The dataset was subjected to the Whale Optimization Algorithm (WOA) for feature selection. The structural threshold extraction is done using the Crow Search Algorithm (CSA). This bio-inspired meta-heuristic algorithm assisted in selecting the most important attributes from the dataset to guarantee that the resulting predictions were as accurate as possible. The choice of optimum hyper parameters affects the prediction results given by a Deep Neural Network (DNN) model. The grid search approach is used to optimize the hyper parameters. The proposed methodology aims to provide a reliable and self-significant system for predicting strokes in early stages. The proposed method is termed as Hybrid Optimization Algorithm (HOA) with reference to pattern segmentation based on attributes and WOA. The HOA tends to apply the schematics of multiple information systems validating and reconfiguring the incoming datasets. The architecture of HOA is presented in Figure 1.

A. Whale Optimization Algorithm

WOA has better performance than recent meta-heuristic methods [15]. In WOA, a population of baleen whales searches for food in a multi-dimensional search space. The individuals' locations are denoted as diverse decision variables. The space among the baleen individual whale and the sustenance relates to the objective cost value. Three operational processes measure the time-dependent position of an individual whale: (1) reduction in encircling prey, (2) bubble-net attacking, and (3) the exploration for prey.

1) Encircling of the Prey

A baleen whale identifies the place of the prey and encompasses it. The optimum design position in the search space is not possible to be known upfront. Hence, the WOA adopts the paramount solution as the target prey closest to the optimum position. Once the best agent is identified, the

remaining agents update their place relatively to the best agent [16]. This behavior is depicted by:

$$\vec{F} = |\vec{D} \cdot \vec{X}_r(a) - \vec{X}(a)| \quad (1)$$

$$\vec{X}(a+1) = \vec{X}_r(a) - \vec{S} \cdot \vec{F} \quad (2)$$

where a indicates the current iteration, \vec{S} and \vec{D} are coefficient vectors, \vec{X}_r is the prey position vector, and \vec{X} represents the position vector of a whale. The vectors \vec{S} and \vec{D} are computed as per the following equation:

$$\vec{S} = 2\vec{s} \cdot \vec{p}_1 - \vec{s} \quad (3a)$$

$$\vec{D} = 2 \cdot \vec{p}_2 \quad (3b)$$

The value of \vec{s} is linearly reduced from 2 to 0 for the repetitions and \vec{p}_1 and \vec{p}_2 are random vectors in $[0, 1]$.

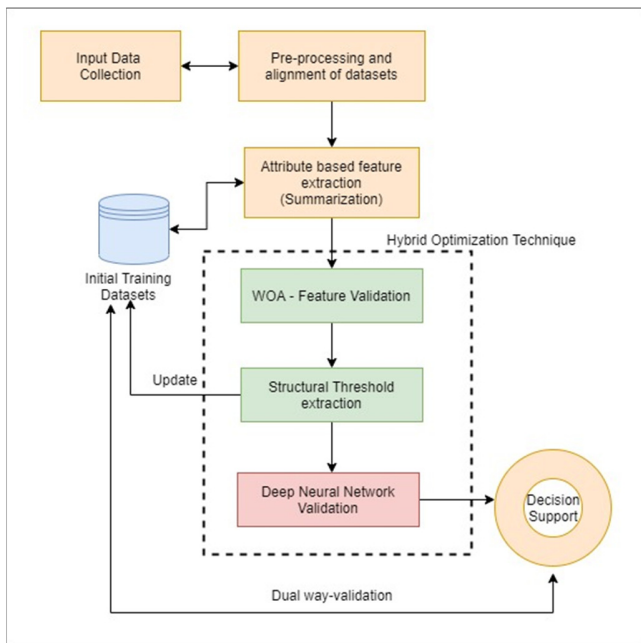


Fig. 1. Block diagram of the proposed HOA.

2) Bubble-net Attacking Method

Two types of approaches are used to represent the bubble-neck behavior of the baleen whales mathematically [17]. These approaches are:

a) Shrinking Encircling Mechanism

The value of \vec{S} is decreased where the variation of \vec{S} is decreased by \vec{s} . A is allocated with an arbitrary value within the interim $[-x, x]$, whereas the value of s is declined from 2 to 0 during repetitions. The new point of the search agent is assigned amongst the original position and the existing agent position by providing arbitrary values for S in $[-1, 1]$.

b) Spiral Updating Position

Finds the space between the whale (A,B) and the prey (A*, B*). The helix-shaped movement is represented using the spiral equation (4):

$$\vec{X}(x+1) = \vec{F}^t e^{yt} \cos(2\pi t) + \vec{X}^*(x) \quad (4)$$

where $\vec{F}^t = |\vec{X}^*(x) - \vec{X}(x)|$ indicates the best solution, y is a constant describing the logarithmic spiral shape, and a is a random value ranging between $[-1, 1]$.

3) The Search Process for the Prey

The vector \vec{S} is varied to search the prey where random whales search randomly based on each other's positions [18]. The mathematical model is:

$$\vec{F} = |\vec{D}X_{rand}^{\rightarrow} - \vec{X}| \quad (5)$$

$$\vec{X}(t+1) = X_{rand}^{\rightarrow} - \vec{S}\vec{F} \quad (6)$$

where X_{rand}^{\rightarrow} is the random position vector of the whale.

B. Crow Search Algorithm

The CSA is based on crow behavior and social interaction. It is a population-based method. Crows are smart birds that live in clusters and have a large brain compared to their body size. They can hide food and memorize locations. Even after several months, they can retrieve the stashed food [19]. The CSA imitates the crow behavior of hiding and recovering food. A population-based algorithm is used. N individuals (crows) make up the flock's size, which is m -dimensional with m being the problem's dimension. The point $X_{x,y}$ of the crow a in a specific iteration y is described as:

$$X_{x,y} = [x_{x,y}^1, x_{x,y}^2, \dots, x_{x,y}^m] \quad (7)$$

where $x=1,2,\dots,N$, $y=1,2,\dots,N$, and N is the number of maximum iterations. Every crow can recall the location Nx,y to hide food until the present iteration.

$$N_{x,y} = [m_{x,y}^1, m_{x,y}^2, \dots, m_{x,y}^n] \quad (8)$$

C. Deep Neural Network

A DNN is a complex neural network with more than two layers. To process data, these networks employ sophisticated mathematical models. A neural network mimics the human brain activity [20]. The network is designed to simulate pattern recognition techniques, with multiple layers of network connections simulating input processing. An input layer, an output layer, and hidden layers in between make up the basic framework of a DNN. Each layer performs its own sorting and ranking, resulting in a feature hierarchy [21]. This advanced neural network aids in processing unstructured and unlabeled data, where artificial intelligence is utilized to classify and organize data outside of standard input-output protocols. The use of DNN in designing health care applications based on Internet of Things technology is studied in [27]. The system is aided with supported, trained datasets for mapping and operation management. These datasets are termed as reliable sources of information prediction and validation. According to regional information scheduling, the processing paradigm is dependent on repeated updating and synchronization of trained datasets. The DNN framework supports the process of validation under HOA. The primary layer of DNN is proposed under the action of input alignment and missing parameter gapping. The DNN middle layer is supported to build a series of interconnected information systems to support reliable

decision support. The overall system is supported under HOA for improvized performance estimation and processing.

IV. EXPERIMENTAL SETUP AND RESULTS

This section discusses the proposed work's experimental findings. Google Colab, was used to implement the proposed task. On the cardiac stroke dataset, Python 3.7 was used to train

and evaluate the proposed model. Missing values were replaced with attribute mean in the pre-processed balanced dataset. The histogram of the dataset is shown in Figure 2 shows the frequency of occurrence of attributes in a certain range of values in continuous and predetermined intervals. The x axis represents the range of values with different scale and the y axis represents the frequency of the value in each attribute.

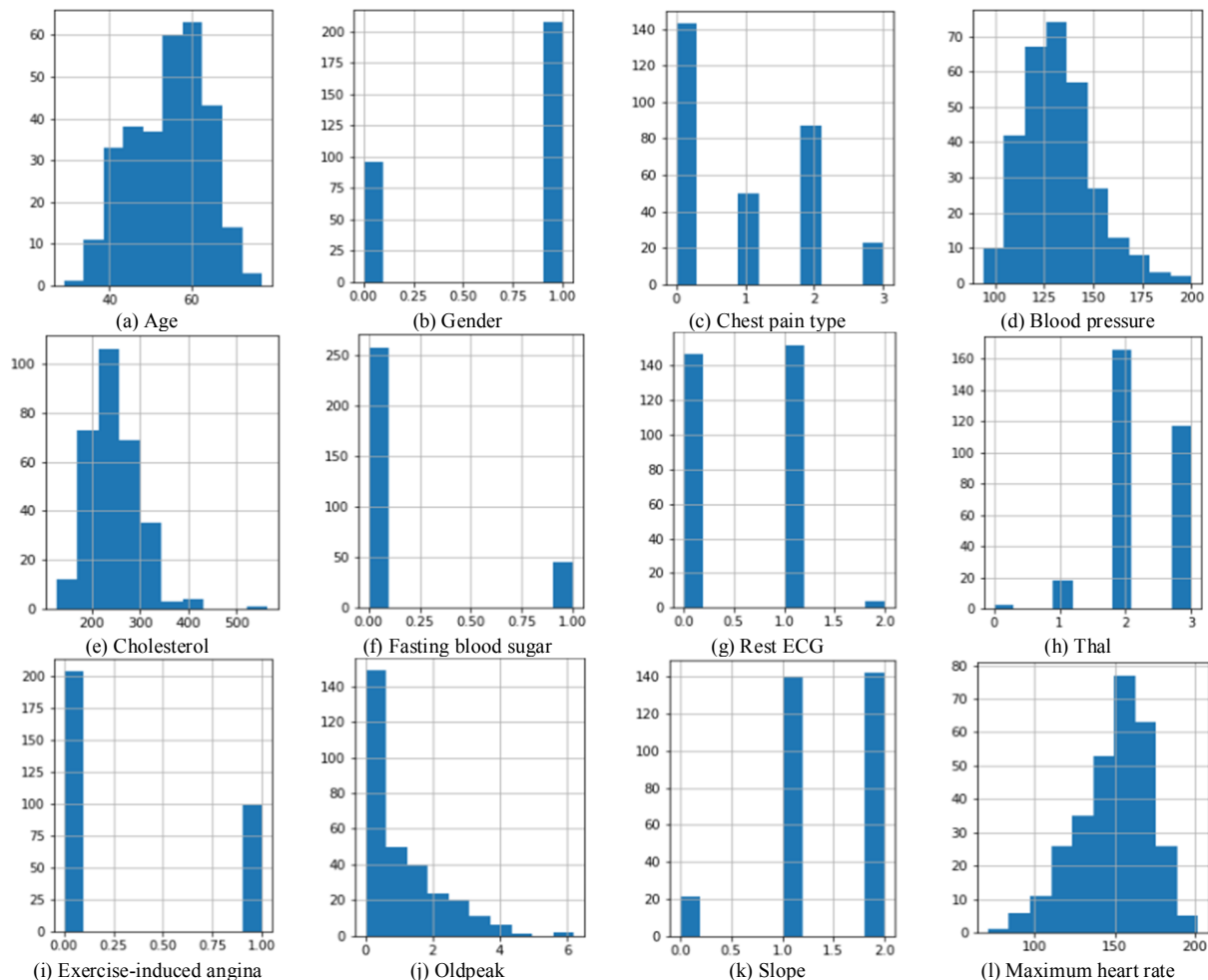


Fig. 2. Histogram of the cardiac stroke dataset.

The heat map is a two-dimensional data visualization where colors represent values, and it is used to understand data easily. Complex datasets can be understood with more sophisticated heat maps. When we want to see which intersections of two lines overlap, a heat map can be informative. The data are more concentrated in categorical values in comparison to others. Figure 3 uses a heat map to depict the correlation between the dataset attributes. Both axes represent the attributes of the dataset and the value in each cell represents the co-relation strength.

WOA and CSA are used to extract interconnected patterns and features. Deep learning algorithms are only capable of working with numerical data. The Label Encoder scheme is used in this study to turn non-numerical into numerical data.

The next stage is to extract features using a hybrid optimization technique. In order to train a deep learning model, it is critical to extract the essential features. By selecting the best attributes that positively impact categorization, feature selection minimizes training time and improves performance.

In this study, a hybrid optimization strategy was employed because it has rapid convergence rate and avoids getting stuck in local minima, which aids in selecting the best features. 70% of the dataset was used to train the model. The proposed model was tested and validated with the remaining 30% of the dataset. The accuracy computation of HOA is significantly higher when compared to previous approaches and techniques in the area of predicting heart strokes and failures. The proposed HOA generated an accuracy of 97.34% surpassing techniques such as

WOA, Ant-Colony Optimization (ACO) algorithm, and DNNs, as presented in Figure 4. The method has also computed stroke prediction rate and validation rate regarding secure computing.

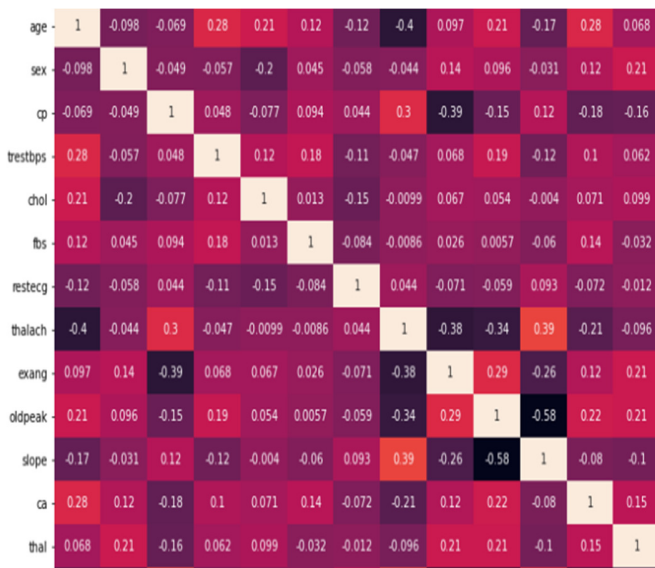


Fig. 3. Heat map of the cardiac stroke dataset.

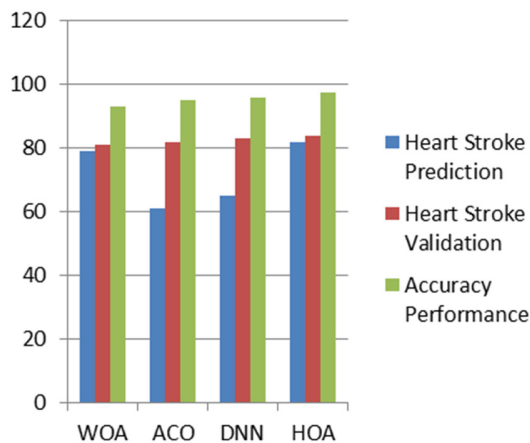


Fig. 4. Comparative analysis on performance efficiency.

TABLE I. COMPARATIVE RESULTS

Ref.	Feature extraction/selection	Classification	Accuracy
[22]	Wavelet transformation	Genetic fuzzy logic system	78.7%
[23]	Laplace smoothing	Naïve Bayes	86%
[11]	-NA-	Hybrid Random Forest	88.7%
[24]	Principal component analysis	Deep Learning	89%
[25]	Fast conditional mutual information	Support Vector Machine	92.37%
Proposed	WOA+CSA	DNN	97.34%

V. CONCLUSIONS AND FUTURE WORK

A deep learning-based framework helps to classify the heart stroke dataset providing valuable insights. The dataset used is the publicly available stroke dataset in [14]. The data were pre-

processed by filling in the missing values and were transformed with the Label Encoder technique for conversion to numerical format. The transformed data were resampled, for treating the imbalances, and normalized using the standard scalar technique. The proposed HOA is based on the schematics of WOA, CSA, and DNN combined in an observed ecosystem of information validation. The proposed HOA exhibited higher prediction rate in comparison with previous techniques and approaches. The trained dataset of HOA is iteratively updated to assure the data threshold is a self-learning framework towards supporting the neural networking model.

ACKNOWLEDGMENT

This research has been funded by the Scientific Research Deanship at University of Ha'il - Saudi Arabia through project number RG-191344.

REFERENCES

- [1] K. Dracup, D. K. Moser, M. Eisenberg, H. Meischke, A. A. Alonzo, and A. Braslow, "Causes of delay in seeking treatment for heart attack symptoms," *Social Science & Medicine*, vol. 40, no. 3, pp. 379–392, Feb. 1995, [https://doi.org/10.1016/0277-9536\(94\)00278-2](https://doi.org/10.1016/0277-9536(94)00278-2).
- [2] *World Heart Federation*. <https://world-heart-federation.org/> (accessed Jul. 20, 2021).
- [3] "Heart Attack and Stroke Symptoms," www.heart.org. <https://www.heart.org/en/about-us/heart-attack-and-stroke-symptoms> (accessed Jul. 20, 2021).
- [4] A. De Vos and N. Soens, *The power of career counseling for enhanced talent and knowledge management*. Northampton, MA, USA: Edward Elgar Publishing, 2008, pp. 119–138.
- [5] C. Jegan, "Heart Attack Prediction System Using Fuzzy C Means Classifier," *IOSR Journal of Computer Engineering*, vol. 14, no. 2, pp. 23–31, Jan. 2013, <https://doi.org/10.9790/0661-1422331>.
- [6] A. H. Chen, S. Y. Huang, P. S. Hong, C. H. Cheng, and E. J. Lin, "HDPS: Heart disease prediction system," in *2011 Computing in Cardiology*, Hangzhou, China, Sep. 2011, pp. 557–560.
- [7] C. S. Dangare and S. S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques," *International Journal of Computer Applications*, vol. 47, no. 10, pp. 44–48, 2012.
- [8] L. Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm," *International Journal of Biology and Life Sciences*, vol. 3, no. 3, pp. 157–160, Jan. 2007.
- [9] F. Ali *et al.*, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, pp. 208–222, Nov. 2020, <https://doi.org/10.1016/j.inffus.2020.06.008>.
- [10] Y. Pan, M. Fu, B. Cheng, X. Tao, and J. Guo, "Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform," *IEEE Access*, vol. 8, pp. 189503–189512, 2020, <https://doi.org/10.1109/ACCESS.2020.3026214>.
- [11] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, <https://doi.org/10.1109/ACCESS.2019.2923707>.
- [12] G. Manogaran, R. Varatharajan, and M. K. Priyan, "Hybrid Recommendation System for Heart Disease Diagnosis based on Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4379–4399, Feb. 2018, <https://doi.org/10.1007/s11042-017-5515-y>.
- [13] S. U. Amin, K. Agarwal, and R. Beg, "Genetic neural network based data mining in prediction of heart disease using risk factors," in *IEEE Conference on Information Communication Technologies*, Thuckalay, India, Apr. 2013, pp. 1227–1231, <https://doi.org/10.1109/CICT.2013.6558288>.

- [14] "UCI Machine Learning Repository: Heart Disease Data Set." <https://archive.ics.uci.edu/ml/datasets/heart+disease> (accessed Jul. 20, 2021).
- [15] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, May 2016, <https://doi.org/10.1016/j.advengsoft.2016.01.008>.
- [16] N. Rana, M. S. A. Latiff, S. M. Abdulhamid, and H. Chiroma, "Whale optimization algorithm: a systematic review of contemporary applications, modifications and developments," *Neural Computing and Applications*, vol. 32, no. 20, pp. 16245–16277, Oct. 2020, <https://doi.org/10.1007/s00521-020-04849-z>.
- [17] G. I. Sayed, A. Darwish, A. E. Hassanien, and J.-S. Pan, "Breast Cancer Diagnosis Approach Based on Meta-Heuristic Optimization Algorithm Inspired by the Bubble-Net Hunting Strategy of Whales," in *Genetic and Evolutionary Computing*, Fuzhou, China, Nov. 2016, pp. 306–313, https://doi.org/10.1007/978-3-319-48490-7_36.
- [18] M. M. Mafarja and S. Mirjalili, "Hybrid Whale Optimization Algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017, <https://doi.org/10.1016/j.neucom.2017.04.053>.
- [19] F. Zhu, D. R. Gonzalez, T. Carpenter, M. Atkinson, and J. Wardlaw, "Lesion Area Detection Using Source Image Correlation Coefficient for CT Perfusion Imaging," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 5, pp. 950–958, Sep. 2013, <https://doi.org/10.1109/JBHI.2013.2253785>.
- [20] G. Montavon, W. Samek, and K.-R. Muller, "Methods for interpreting and understanding deep neural networks," *Digital Signal Processing*, vol. 73, pp. 1–15, Feb. 2018, <https://doi.org/10.1016/j.dsp.2017.10.011>.
- [21] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2012, pp. 3642–3649, <https://doi.org/10.1109/CVPR.2012.6248110>.
- [22] T. Nguyen, A. Khosravi, D. Creighton, and S. Nahavandi, "Classification of healthcare data using genetic fuzzy logic system and wavelets," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2184–2197, Mar. 2015, <https://doi.org/10.1016/j.eswa.2014.10.027>.
- [23] V. Cherian and M. S. Bindu, "Heart Disease Prediction Using Naïve Bayes Algorithm and Laplace Smoothing Technique," *International Journal of Computer Science Trends and Technology*, vol. 5, no. 2, pp. 68–73, 2017.
- [24] S. Tuli *et al.*, "HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments," *Future Generation Computer Systems*, vol. 104, pp. 187–200, Mar. 2020, <https://doi.org/10.1016/j.future.2019.10.043>.
- [25] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020, <https://doi.org/10.1109/ACCESS.2020.3001149>.
- [26] N. K. Al-Shammari, H. B. Almansour, and M. B. Syed, "Development of an Automatic Contactless Thermometer Alert System Based on GPS and Population Density," *Engineering, Technology & Applied Science Research*, vol. 11, no. 2, pp. 7006–7010, Apr. 2021, <https://doi.org/10.48084/etasr.4103>.
- [27] N. K. Al-Shammari, T. H. Syed, and M. B. Syed, "An Edge – IoT Framework and Prototype based on Blockchain for Smart Healthcare Applications," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7326–7331, Aug. 2021, <https://doi.org/10.48084/etasr.4245>.